AD-A086 203 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
NONPARAMETRIC BAYESIAN ESTIMATION OF THE HORIZONTAL DISTANCE BE--ETC(U)
MAR 80 M HOLLANDER, R KORWAR AFOSR-78-3678
UNCLASSIFIED FSU-STATISTICS-M537 AFOSR-TR-80-0461 NL

1 OF 1
AD
A086203

END
DATE
FILMED
8-80
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

12

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-0461	2. GOVT ACCESSION NO.: AD-A086 203	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): NONPARAMETRIC BAYESIAN ESTIMATION OF THE HORIZONTAL DISTANCE BETWEEN TWO POPULATIONS		5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Myles Hollander, Ramesh Korwar		8. CONTRACT OR GRANT NUMBER(s): AFOSR-78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Florida State University, Department of Statistics, Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE: March 1980
		13. NUMBER OF PAGES: 10
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 11		15. SECURITY CLASS. (of this report): UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

FSU-STATISTICS-M537

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

TR-78-104

DTIC ELECTE JUL 0 7 1980 S D

E

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Horizontal distance between two distributions, Nonparametric estimator, Bayesian estimator, Dirichlet process

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The horizontal distance $\Delta(x) = G^{-1}(F(x)) - x$ has been shown by Doksum (1974) to be a useful measure of the difference, at each x, between the populations defined by continuous distribution functions F(x) and G(x). Here we assume that G is know, and we develop a Bayesian nonparametric estimator $\hat{\Delta}_n(x)$ of $\Delta(x)$ based on a random sample of n X's from F. The estimator $\hat{\Delta}_n$ is, for weighted squared-error loss, Bayes with respect to Ferguson's (1973) Dirichlet process prior. Using a result of Korwar and Hollander (1976), the Bayes risk of $\hat{\Delta}_n$ is evaluated for the case when G is uniform.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 400 277

ADA086203

DDC FILE COPY

AFOSR-TR- 80-0461

NONPARAMETRIC BAYESIAN ESTIMATION OF THE HORIZONTAL DISTANCE BETWEEN TWO POPULATIONS

by

Myles Hollander[1] and Ramesh Korwar[2]

The Florida State University
and
University of Massachusetts

FSU Statistics Report M537
AFOSR Technical Report No. 78-104

March, 1980
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

[1]Research supported by the Air Force Office of Scientific Research, AFSC, USAF under Grant AFOSR-78-3678.

[2]Research supported by the Air Force Office of Scientific Research, AFSC, USAF under Grant F49620-79-C-0105.

80 7 2 112

ABSTRACT

The horizontal distance $\overset{\text{delta}}{\Delta}(x) = G^{-1}(F(x)) - x$ has been shown by Doksum (1974) to be a useful measure of the difference, at each x, between the populations defined by continuous distribution functions F(x) and G(x). Here we assume that G is known, and we develop a Bayesian nonparametric estimator $\hat{\Delta}_n(x)$ of $\Delta(x)$ based on a random sample of n X's from F. The estimator $\hat{\Delta}_n$ is, for weighted squared-error loss, Bayes with respect to Ferguson's (1973) Dirichlet process prior. Using a result of Korwar and Hollander (1976), the Bayes risk of $\hat{\Delta}_n$ is evaluated for the case when G is uniform.

1. Introduction.

When F and G are continuous distribution functions, the horizontal distance

$$\Delta(x) = G^{-1}(F(x)) - x, \ x \text{ real}, \tag{1}$$

has been shown by Doksum (1974) to be a useful measure of the difference, at each x, between F and G. Under suitable regularity, Doksum shows that $\Delta(x)$ is essentially the only function satisfying

$$X + \Delta(X) \overset{d}{=} Y, \tag{2}$$

where, in (2), X is distributed according to F, Y is distributed according to G, and "$\overset{d}{=}$" means "has the same distribution as."

When the linear model

$$F(x) = G(x + \Delta), \text{ for all } x, \tag{3}$$

holds, where Δ is a constant, then $\Delta(x) \equiv \Delta$ (and, of course, when $F \equiv G$, $\Delta(x) \equiv 0$.)

When one observes a random sample of n X's from F and an independent random sample of m Y's from G, Doksum suggests estimating $\Delta(x)$ by

$$\hat{\Delta}_N(x) = G_m^{-1}(F_n(x)) - x, \tag{4}$$

where $N = m + n$ and F_n, G_m are the empirical distribution functions based on the X's and Y's, respectively. Doksum also derives a simultaneous confidence band for $\Delta(x)$ and shows that $N^{\frac{1}{2}}\{\hat{\Delta}_N(x) - \Delta(x)\}$ converges weakly to a Gaussian process.

In this paper we consider the one-sample problem where *G is known* and (just) a random sample of n X's from F is available for estimating $\Delta(x)$. One natural estimator for this problem is the one-sample limit ($m \to \infty$) of Doksum's estimator $\hat{\Delta}_N$. This one-sample limit is

$$\hat{\Delta}_n(x) = G^{-1}(F_n(x)) - x. \tag{5}$$

The estimator $\hat{\Delta}_n$ does not utilize prior information about the unknown F. Our approach is Bayesian and leads to an estimator $\tilde{\Delta}_n$ which does use prior information about F.

We assume that F is a *random* distribution function chosen according to Ferguson's (1973) Dirichlet process prior (Definition 2.2) with parameter $\alpha(\cdot)$, a completely specified measure on the real line R with the Borel σ-field B. A defect to this approach is that the randomly chosen F will not be continuous (Ferguson's Dirichlet process prior chooses, with probability one, a discrete distribution) and thus the desirability of estimating $\Delta(x)$ is slightly diminished. Nevertheless, in this case $\Delta(x)$ remains a useful measure of the distance between F and G at x, and the resulting estimator $\tilde{\Delta}_n(x)$ combines sample information and prior information in an effective manner.

Our loss function is

$$L(\hat{\Delta}, \Delta) = \int (\hat{\Delta}(x) - \Delta(x))^2 dW(x), \tag{6}$$

where $\hat{\Delta}$ is an estimator of Δ and W is a finite measure on (R, B). A general expression for the Bayes estimator $\tilde{\Delta}_n$ is given in Section 3, and explicit expressions for $\tilde{\Delta}_n$ are obtained for the cases when G is (i) exponential and (ii) uniform. Furthermore, in the uniform case we derive the Bayes risk of $\tilde{\Delta}_n$.

Section 2 contains preliminaries relating to the Dirichlet process.

2. Dirichlet Process Preliminaries.

This section briefly gives some definitions and theorems associated with the Dirichlet process. For further details the reader is refered to Ferguson (1973).

DEFINITION 2.1 (Ferguson). Let $Z_1, \ldots, Z_k$ be independent random variables with Z_j having a gamma distribution with shape parameter $\alpha_j \geq 0$ and scale parameter 1, $j = 1, \ldots, k$. Let $\alpha_j > 0$ for some j. The ***Dirichlet distribution*** with parameter $(\alpha_1, \ldots, \alpha_k)$, denoted by $D(\alpha_1, \ldots, \alpha_k)$, is defined as the distribution of $(Y_1, \ldots, Y_k)$, where $Y_j = Z_j / \sum_{i=1}^{k} Z_i$, $j = 1, \ldots, k$.

Since $\sum_{i=1}^{k} Y_i = 1$, the Dirichlet distribution is singular with respect to Lebesgue measure in k-dimensional space. If $\alpha_j = 0$, the corresponding Y_j is degenerate at zero. If however $\alpha_j > 0$ for all j, the $(k - 1)$-dimensional distribution of $(Y_1, \ldots, Y_{k-1})$ is absolutely continuous with density

$$f(y_1, \ldots, y_{k-1} | \alpha_1, \ldots, \alpha_k) \tag{7}$$

$$= \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_k)} \left(\prod_{j=1}^{k-1} y_j^{\alpha_j - 1}\right)\left(1 - \sum_{j=1}^{k-1} y_j\right)^{\alpha_k - 1} I_S(y_1, \ldots, y_{k-1})$$

where S is the simplex $S = \{(y_1, \ldots, y_{k-1}): \ y_j \geq 0, \sum_{j=1}^{k-1} y_j \leq 1\}$.

DEFINITION 2.2 (Ferguson). Let (X, A) be a measurable space. Let α be a non-null finite measure (nonnegative and finitely additive) on (X, A). We say P is a ***Dirichlet process*** on (X, A) with parameter α if for every $k = 1, 2, \ldots$, and measurable partition $(B_1, \ldots, B_k)$ of X, the distribution of $(P(B_1), \ldots, P(B_k))$ is Dirichlet with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$.

DEFINITION 2.3 (Ferguson). The $\mathcal{X}$-valued random variables $X_1, \ldots, X_n$ constitute *a sample of size n from a Dirichlet process P* on $(\mathcal{X}, \mathcal{A})$ with parameter α if for any $m = 1, 2, \ldots$ and measurable sets $A_1, \ldots, A_m, C_1, \ldots, C_n$, $Q\{X_1 \in C_1, \ldots, X_n \in C_n | P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n)\} = \prod_{i=1}^{n} P(C_i)$ a.s., where Q denotes probability.

THEOREM 2.4 (Ferguson). *Let P be a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter* α, *and let $X_1, \ldots, X_n$ be a sample of size n from P. Then the conditional distributioon of P given $X_1, \ldots, X_n$ is a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter* $\beta = \alpha + \sum_{i=1}^{n} \delta_{X_i}$, *where, for* $x \in \mathcal{X}$, $A \in \mathcal{A}$, $\delta_x(A) = 1$ *if* $x \in A$, 0 *otherwise.*

3. A Bayes Estimator of the Horizontal Distance.

We suppose that F is chosen according to a Dirichlet process prior on (R, B) with parameter α. With the loss function given by (6), the Bayes estimator for the no-sample problem is found by minimizing the right-hand-side of (8),

$$EL(\tilde{\Delta}, \Delta) = \int E(\tilde{\Delta}(x) - \Delta(x))^2 dW(x), \tag{8}$$

where the expectation is with respect to F. The estimator is obtained by minimizing $E(\tilde{\Delta}(x) - \Delta(x))^2$ for each x, yielding

$$\tilde{\Delta}(x) = E(\Delta(x)) = E\{G^{-1}F(x)\} - x. \tag{9}$$

We next evaluate (9) in the cases where (i) G is exponential and (ii) G is uniform.

3.1. The Case Where G is Exponential: Let $G(x) = 1 - \exp(-\lambda x)$, $x > 0$, and 0 for $x \le 0$, for some $\lambda > 0$. Then

$$G^{-1}(x) = -\lambda^{-1} \ln(1 - x), \quad 0 < x < 1,$$

and (9) reduces to

$$\tilde{\Delta}(x) = \{B(\alpha', \beta')\}^{-1} \int_0^1 [-\lambda^{-1} \ln(1 - y)] y^{\alpha'-1}(1 - y)^{\beta'-1} dy \quad - x, \tag{10}$$

where $B(\alpha', \beta') = \Gamma(\alpha')\Gamma(\beta')/\Gamma(\alpha' + \beta')$. Equation (10) makes use of the fact that for each x, F(x) is distributed according to the Beta distribution with parameters $\alpha' = \alpha((-\infty, x])$, $\beta' = \alpha(R) - \alpha'$. (To see this use Definition 2.2 with the measurable partition $B_1 = (-\infty, x]$, $B_2 = R - B_1$.) Thus, for the "no-sample" problem, by expanding $\ell n(1 - y)$ in a power series, we obtain

$$\tilde{\Delta}(x) = \{\lambda B(\alpha', \beta')\}^{-1} \int_0^1 \sum_{j=1}^{\infty} j^{-1} y^{\alpha'+j-1}(1 - y)^{\beta'-1} dy \quad - x$$

$$= \lambda^{-1} \sum_{j=1}^{\infty} [B(\alpha' + j, \beta')/\{jB(\alpha', \beta')\}] \quad - x.$$

Using Theorem 2.4, the Bayes estimator when a sample $X_1, \ldots, X_n$ is available from F, is

$$\tilde{\Delta}_n(x) = \lambda^{-1} \sum_{j=1}^{\infty} [B(\alpha'' + j, \beta'')/\{jB(\alpha'', \beta'')\}] \quad - x, \qquad (11)$$

where

$$\alpha'' = \alpha((-\infty, x]) + \sum_{i=1}^{n} \delta_{X_i}((-\infty, x]),$$

$$\beta'' = \alpha(R) + n - \alpha''.$$

3.2. The Case Where G is Uniform: Let $G(x) = 0$ for $x < a$, $(x - a)/(b - a)$ for $a \le x \le b$, and 1 for $x > b$, for some $a < b$. Then (9) reduces to

$$\tilde{\Delta}(x) = \int_0^1 [y(b - a) + a]\{B(\alpha', \beta')\}^{-1} y^{\alpha'-1}(1 - y)^{\beta'-1} dy \quad - x$$

$$= a + (b - a)\{B(\alpha' + 1, \beta')/B(\alpha', \beta')\} \quad - x$$

$$= a + (b - a)\{\alpha'/(\alpha' + \beta')\} \quad - x$$

$$= a + (b - a)F_0(x) \quad - x,$$

where

$$F_0(x) = \alpha((-\infty, x])/\alpha(R), \; x \in R,$$

can be interpreted as the "prior guess" at F.

Thus, from Theorem 2.4, when a sample $X_1, \ldots, X_n$ is available from F, the Bayes estimator is

$$\tilde{\Delta}_n(x) = a + (b - a)\tilde{F}_n(x) - x, \ x \in R, \tag{12}$$

where

$$\tilde{F}_n(x) = \{\alpha((-\infty, x]) + \sum_{i=1}^{n} \delta_{X_i}((-\infty, x])\}/\{\alpha(R) + n\}.$$

The minimum Bayes risk $S(\alpha)$ of $\tilde{\Delta}_n$(12) can be computed using results of Korwar and Hollander (1976). Korwar and Hollander obtained the minimum Bayes risk $R(\alpha)$ of the estimator $\tilde{F}_n$ against weighted squared error loss to be

$$R(\alpha) = [\alpha(R)/\{(\alpha(R) + 1)(\alpha(R) + n)\}]\int F_0(x)(1 - F_0(x))\,dW(x).$$

(See equation (2.19) of Hollander and Korwar (1976) and replace the m of that equation with n here.) It immediately follows that $S(\alpha) = (b - a)^2 R(\alpha)$.

We note that we can also directly obtain the risk $T(\alpha)$ (say) of the one sample limit of Doksum's estimator $\hat{\Delta}_n$ (see equation (5)) with respect to the Dirichlet process prior with parameter α in this case when G is uniform. We find

$$\hat{\Delta}_n(x) = a + (b - a)F_n(x) - x, \ x \in R,$$

where $F_n(x)$ is the empirical distribution function of the X's. Using (3.3) of Korwar and Hollander (1976) we obtain $T(\alpha) = (b - a)^2(1 + \alpha(R)/n)R(\alpha)$.

REFERENCES

[1] Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. Ann. Statist. 2, 267-277.

[2] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209-230.

[3] Korwar, R. M., and Hollander, M. (1976). Empirical Bayes estimation of a distribution function. Ann. Statist. 4, 581-588.

ILMED

8